Users Dealing with Spam and Spam Filters: Some Observations and Recommendations[1]

Christopher Lueg and **Sam Martin**
School of Computing
University of Tasmania
Private Bag 100
Hobart TAS 7005, Australia
christopher.lueg@utas.edu.au

ABSTRACT
The email communication system is threatened by unsolicited commercial email aka spam. In response, spam filters have been deployed widely to help reduce the amount of spam users have to cope with. This paper describes work towards helping users better understand the often complex decision making that is spam filtering. An investigation of a number of popular web-based email services suggests that the filtering process is typically implemented as a black box allowing very little user involvement. In order to explore how we could help users understand how spam filters work and how they assess messages we conducted a number of user experiments using a simulated email interface providing richer spam filtering information than the webmail interfaces we investigated. Feedback indicates that additional information provided by the interface would be welcome and suggests to further investigate ways to involve users in the filtering process.

Author Keywords
Electronic mail, Information search, Information retrieval, Informative Interfaces, Spam.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Usability and thus usefulness of email communication is threatened by large amounts of "spam" which is a colloquial term denoting unsolicited commercial email or bulk email (short: UCE or UBE, respectively). Spam has been called "an epidemic [...] that we need to learn how to control" by the acting chief of the Australian Communications Authority, Robert Horton (reported by Cage 2004). There are predictions that spam is threatening to destroy email as a way of communicating (e.g., Whitworth and Whitworth 2004). There is an abundance of papers exploring technical ways to address the spam filtering problem (e.g., NOIE 2002; CEAS is a new international conference series focusing on email and spam filtering). Among others, Li (2006) discusses legislative ways to address the problem.

Spam filtering remains a challenge. Advanced filtering approaches can be considered fairly reliable but, at the same time, there is also considerable anecdotal evidence that overly ambitious spam filters cause problems for genuine emails (eg Lueg 2004; Lueg et al. 2007). Problems appear to be related to the lack of *objective* criteria that spam filters could employ for determining "solicitedness" of emails. A core problem is that (un)solicitedness is the defining characteristic of spam. The Australian National Office for the Information Economy (NOIE), for example, defined spam as "unsolicited electronic messaging, regardless of its content" (NOIE 2002, p. 7). Similarly, the U.S. Center for Democracy and Technology, one of a number of sources providing equivalent definitions, states that "[the term] spam is used to refer to a single or multiple pieces of mail that are perceived by the recipients to be unsolicited and unwanted" (CDT n.d.). In other words, the problem that spam filters are facing is that neither "unsolicited" nor "unwanted" are objective, measurable aspects of emails (Lueg 2005).

The likely persistence of spam filtering problems in the foreseeable future suggests to pay more attention to improving how users can handle and influence spam filtering results. A brief literature review indicated a distinct

1Final version published in the Proceedings of the 8th Annual ACM SIGCHI-NZ Conference on Human-Computer Interaction, Hamilton, New Zealand, 1-4 July 2007, pp. 67-72.

lack of papers investigating the user's role in dealing with spam and how he or she can influence spam filtering outcomes. This comes as a surprise as after all, spam filtering is about supporting users in coping with spam. Based on interviews conducted in the U.S., Fallows (2003) even suggests that 30% of email users surveyed were concerned their email filters might filter genuine incoming email and 23% of users were concerned email they send to others may be filtered.

In this paper we describe work towards better understanding the spam filtering process, from an HCI point of view. First, we look at popular web-based email (webmail) services and explore what information their interfaces offer in order to help users understand how the built-in spam filters work and how they can be influenced. Then, we report on user experiments we conducted using a simulated email and spam filter interface. The interface provides richer spam filtering information than the webmail interfaces we investigated. Next, we discuss the feedback we received. The papers closes with a discussion of the findings and an outlook on future research in this increasingly important area.

PART I: SUPPORT CURRENTLY OFFERED BY WEBMAIL

Scope of the Investigation

We investigated the user interfaces provided by a number of web-based email services ("webmail"). The objective was to determine how they support users in dealing with the outcomes of spam filtering processes, such as spam messages not recognized as such (false negatives) or genuine email classified as spam (false positives).

We focused on "free" webmail services even though these services have a reputation of being "throw away" accounts. At the same time there is ample anecdotal evidence suggesting that free webmail accounts are used for genuine email purposes (e.g. "At work around 20% of legitmate[sic] non-spam emails , mostly inquiries, are from Hotmail or other webmail services." http://www.oreillynet.com/cs/user/view/cs_msg/38427). Other anecdotal evidence includes that in a recent university course half of the about 50 students provided free webmail addresses (hotmail.com, yahoo.com, yahoo.com.au, gmail.com) as part of their official contact details.

The webmail services we investigated were hotmail.com, yahoo.com.au, gmail.com and gmx.net. All services expect users to log on to their respective web sites for sending and receiving email even though some services also support POP3/IMAP mailbox access for users preferring to use their own email clients, such as Outlook Express or Eudora.

We investigated the webmail services using computers connected to Australian Internet service providers (ISPs). Users accessing the same services from other countries may be presented different information and/or features due to service customization.

Information about the spam filtering process

In this section we look at the information provided by the chosen webmail interfaces in order to help users understand what the built-in spam filters are doing and why. This includes, in particular, information about spam definitions and/or spam characteristics employed by spam filters when determining 'spamminess'.[2]

Information provided by hotmail.com

Hotmail offers customers to "*stop receiving junk e-mail (spam)"* but the site's online help does not provide a specific definition as to what the webmail provider actually considers to be "junk e-mail" or "spam". The junk filter setup window merely mentions that the filter *"helps keep unsolicited messages out of your Inbox."* The spam filter can be set to different levels which are referred to as *Low*, *Enhanced* and *High*. It does not seem to be possible to turn the spam filter *off*.

Information provided by yahoo.com.au

Yahoo7 offers customers to *"Customise our anti-spam tools to maximise your spam protection"*. Clicking the *Spam Protection* link invites customers to adjust what *SpamGuard* does to messages identified as Spam: either delete these messages upon receipt without opportunity to view them or save the messages in the Bulk Folder for a period of up to one month.

Following the *What's this?* link next to *Turn SpamGuard OFF* leads to more information about the Bulk Mail Folder. In the early stages of the SpamGuard setup process, yahoo does not appear to distinguish between spam and bulk email. In a section on *"What is the difference between solicited and unsolicited commercial email?"* yahoo explains the situation in more detail though: *"[u]nsolicited commercial email, commonly known as spam, is any message or posting, regardless of its content, that is sent to multiple recipients who have not specifically requested the mail. Solicited commercial email is any commercial message, newsletter, or posting sent only to recipients who have requested it and can choose to opt out of receiving the mailing."* Yahoo's online help provides further information regarding their (informal) definition of spam: *"Spam is any message or posting, regardless of its content, that is sent to multiple recipients who have not specifically requested the mail. It can also be multiple postings of the same message to newsgroups or list servers that aren't related to the topic of the message. Other common terms for spam include UCE (Unsolicited Commercial Email) and UBE (Unsolicited Bulk Email)."* (http://help.yahoo.com/help/au/mail/spam/spam-02.html) Customers who do not consider this information sufficient have the choice of contacting Yahoo's *Customer Care*. Apart from these rather

2Some argue details of the filtering criteria should not be revealed as the information could help spammers find new ways to bypass spam filters. However, the maintainers of the widely used SpamAssassin spam filter do provide this information on their web site without becoming irrelevant.

informal descriptions, Yahoo does not seem to provide information about the (technical) criteria they use to compute *spamminess* of messages.

Similar to Hotmail, Yahoo acknowledges they *"may occasionally send [solicted bulk messages] to [the user's] bulk mail folder" although Yahoo's intent is "to send solicited emails to [the user's] Inbox"*. If this happens Yahoo request *"to let [them] know by forwarding the message to a Yahoo! Customer Care associate"*.

Information provided by gmail.com

Gmail does not seem to provide any information about their spam filtering activities. Only the fact that they offer a *Spam* folder and the existance of a *Not Spam* button suggests that Gmail is applying some sort of spam filtering. Something the writer initially believed to be a spam message ("French Fry Spam Casserole - Bake 30-40 minutes") turned out to be what Google calls a "Web (advertising) Clip" which was placed right above the email inbox therefore causing some confusion.

Information provided by gmx.net

Webmail provider Global Message Exchange (GMX) state their anti-spam measures are able to reduce the incoming spam load by up to 98%. They mention using seven different anti-spam modules plus a list of reliable email providers ("Trusted Networks") provided by an *Anti Spam Task Force*. GMX's online help provides comprehensive information about the spam situation including information about how allegedly spammers harvest email addresses and advice how to avoid spam. GMX also provide fairly detailed textual information regarding the different spam filtering modules they offer.

Supporting Users in Interacting with the Spam Filtering Process

In the previous section we investigated information provided by webmail interfaces in order to help users understand what respective webmail providers consider spam and how this possibly impacts on their email.

In a nutshell, the information provided is hardly sufficient to understand how spam filters evaluate messages.

The question we address in this section is whether the webmail interfaces provide ways for email users to *influence* current and future spam filtering outcomes. This may be necessary in the case of false positives (genuine email classified as spam) but also in the case of false negatives (spam not recognized as such).

hotmail.com's user support

Hotmail encourages users to set Hotmail's spam filter level to *Low*, *Enhanced* or *Exclusive*. As mentioned earlier it remains unclear what hotmail actually considers *"obvious junk"* as mentioned in the *Low* setting as Hotmail does not seem to provide a concise definition of spam. The structure of the selection menu (see illustration 1) means users are unable to turn spam filtering *off* which is functionally equivalent to Hotmail reserving the right to discard certain emails they consider spam. Anecdotal evidence suggests that Hotmail happens to discard genuine email without warning (eg "So wie es aussieht, wirft Hotmail die Bestätigungs-eMails des Moodle-Systems kommentarlos weg, d.h. sie finden sich nicht einmal im Spamordner. " http://intrepid.interactivesystems.info/moodle/index.php?lang=en_utf8).

Having set the junkmail filter setting to "*Low*" we were repeatedly encouraged to increase the setting to "Enhanced" to reduce the amount of junk e-mail in the inbox even though the account we were using for test purposes received hardly any junk mail.

Illustration 1: hotmail's selector

Hotmail explicitly acknowledges that their spam filters may falsely classify genuine messages as spam: *"at enhanced and exclusive filtering levels, wanted messages are occasionally identified as junk e-mail"*. The only way Hotmail customers may deal with false-positives is checking their junk mail folder regularly (expire time is mere five days).

Hotmail provides ways to report spam by clicking a *"Junk"* (Report as junk e-mail) button but does not reveal how reporting is going to address the filtering issue the user experienced. Hotmail only state that *"[m]essages you report as junk e-mail are used by Microsoft to improve the Junk E-Mail Filter. Microsoft may also provide the reported messages to third parties to help combat junk e-mail."*

Hotmail does not provide information as to why an email was classified as spam or how reporting will influence the future performance of the spam filter.

Yahoo.com.au's user support

If Yahoo customers decide to have SpamGuard turned on, they are encouraged to check the bulk email folders regularly: *"When SpamGuard is ON, Yahoo! Mail will deliver spam to your Bulk folder and periodically delete the messages. You can specify how frequently you would like Yahoo! Mail to delete messages in your Bulk folder."*

Once email has been received and classified as spam, Yahoo introduces the "bulk email folder" by sending a

respective email: *"Yahoo! Mail has just created a Bulk folder for your account [...] While we make our best efforts to deliver solicited commercial and non-commercial email directly to your inbox, a non-spam message may be delivered to your Bulk folder on occasion. For this reason, we recommend that you check this folder periodically to ensure you don't miss important messages. [...]"*.

Users are also encouraged to provide feedback by clicking *"Spam"* and *"Not Spam"* buttons. Similar to Hotmail, Yahoo does not provide information as to why an email was assessed to be spam. Neither do they reveal as to how reporting spam will influence the future spam filter performance.

Support provided by gmail.com

Similar to other services reviewed in this paper, Gmail users are also encouraged to provide feedback by clicking a *Report Spam* button in the case of false negatives or a *Not Spam* button in the case of false positives that were filed in the *Spam* folder.

Gmail does not reveal as to how reporting will influence the future performance of the spam filter either. Neither does Gmail provide information as to why an email was assessed to be spam.

Support provided by gmx.net

GMX keeps reminding users to turn *on* spam filtering if it is not activated. The webmail service is the only service that provides at least some information as to why messages were classified as spam (see illustration 2; "H" means high level spamminess was attributed based on an analysis of the message header; "A" indicates the use of a global black list; "G" refers to GMX's own black list).

	Absender	Betreff	Datum	kB	
H	"Cathryn"	th1s h0t amateur ch1c...	04.07.07 00:28	2	
S	"Magic-Jackpot C...	Bis 1000 Euro frei!	03.04.07 19:18	5	
S	Cleo Mosley	asian slut riding fuc...	03.04.07 15:31	2	
A	"Angie"	Thinking about you	03.04.07 10:57	21	
S	Tara Erickson	SEXUALLY EXPLICIT : c...	03.04.07 10:17	2	
S	"Sharon"	SEXUALLY EXPLICIT : t...	03.04.07 08:36	2	
A	"probe"	Viagra soft for $1.62...	02.04.07 21:34	9	
A	"quite Unlike"	Short 30 second form	03.04.07 11:36	2	
S	Alice Cunningham	RE: Damiana one of t...	03.04.07 05:33	2	
H	Robert Guerrero	Brunette is fond of t...	02.04.07 23:24	2	

Illustration 2: GMX filter information

"Informating" and Informing

In the previous two sections we illustrated to what extent a number of popular webmail interfaces keep their users informed about spam filtering activities. In a nutshell, most interfaces do not provide much information to users; GMX the only exception. This means the process is treated as if spam filtering was merely automated not *informated* (Zuboff 1988) as it arguably is.

The interesting point is that spam filtering systems *are able* to provide users with comprehensive information about what is happening to their emails and therefore *empower* users. As mentioned earlier, an argument could be made that disclosure of too many filtering details would undermine the effectiveness of filters as spammers might use the information to improve their spam strategies. However, SpamAssassin does provide rich information and is still considered one of the best spam filtering systems available. Furthermore, it can be argued that users should be able to find out what impacts the success (of failure) of their communication, via email.

Spam arises from an online social situation that was created through the deployment of communication technology (e.g., Whitworth and Whitworth 2004). This means spam filtering raises broader issues of consent as spam filtering is a way to deal with a particular aspect of a complex socio-technical system. Solutions to the problems created necessarily exceed the technical realm as neither "unsolicited" nor "unwanted" are objective, measurable aspects of emails (Lueg 2005). Whitworth and Whitworth (2004) focus on legitimacy of the communication process (and ways to fight communication they consider illegitimate) whereas we are more interested in *legitimacy* and the issue that *legitimacy* is not an objective, measurable aspect of messages.

This *crux* means users and spam filters operate in different ontological spaces and mediating between these spaces requires human-oriented development processes (e.g., Norman 1998; Preece et al. 2002) producing technology that supports users, not vice versa.

European webmail provider GMX features the highest degree of *informing* as users are offered a range of options including activating and deactivating certain spam filtering modules. The options that are available are explained in reasonable detail.

The spam filtering process appears to be fully transparent but in fact the very details of the filtering process are not disclosed to the user. The interface lists modules that were responsible for classification of messages as spam but most likely several modules are involved in assessing *spamminess*. The information provided by the GMX interface should probably be seen as providing the dominant contributing factor for concluding that a message is spam.

PART II: USER PERCEPTIONS REGARDING AN INFORMATIVE EMAIL INTERFACE

In order to find out more about user perceptions regarding the information provided by spam filters we conducted a number of user experiments using a simulated email interface providing richer spam filtering information than most of the web-based email interfaces discussed in the previous sections (see Martin 2006 for details). The idea was to find out if users would actually appreciate the additional information.

Scope of the Investigation

Subjects for evaluating the interface were recruited among

Computing students, via email. In total we had 15 participants who evaluated the interface for about 20 minutes each (14 male, 1 female; age range 20-47, most early-mid 20ies; most were core computing students but some were doing combined degrees including information systems). Prior to the experiments participants were briefed and received information about the ethics approval we obtained prior to preparing the experiments.

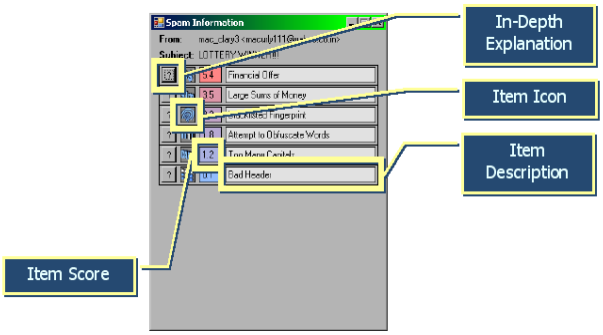
During the experiments the email interface simulated a number of realistic email scenarios by providing a mix of actual spam messages and custom-made emails. These scenarios included an "introductory" scenario featuring correctly classified genuine messages and spam messages. More demanding scenarios included incorrectly classified emails including false positive and false negatives.

Two of the more advanced scenarios covered finding out why the spam filter incorrectly classified certain *incoming* messages and determining why the interface warned that *outgoing* emails might be classified as spam by the recipient's spam filter.

Key spam-related aspects of the interface included:

1. A *spam icon* indicating the dominant reason for classifying the email as spam;
2. A *spam bar* reflecting the degree of spamminess thus providing more detailed information regarding the spam status;
3. Upon request the specific *spam score* the email attracted;
4. Subtle *highlighting* of emails assessed to be spam

Upon opening a specific message the interface offered detailed *Spam Information* (see picture) revealing all the different aspects of the message that were seen as *contributing* to *spamminess*.



By contrast, the most informative of the interfaces discussed above (GMX) would only indicate what is believed to be the dominant contributing factor. GMX would neither reveal exact scores nor further contributing factors.

Outcomes of the Investigation

The evaluation incorporated 35 multiple-choice questions and 2 open feedback questions, all to be addressed right after the actual experiment. 19 of the questions were related to the user experience, 16 questions addressed program usage issues.

Most subjects rated their email experience (1=not at all; 5 =very) as high (Mean 4.40; Std Dev 0.83); spam experience was also rated as high (Mean 4.27; Std Dev 1.03). However the experience with spam filtering as well as spam filtering process knowledge (1=none; 5=a lot) was rated at mere 2.80 (Std Dev 1.08) and 2.79 (Std Dev 1.05), respectively.

Regarding filter control and awareness it is interesting to note that 7 subjects (46.5%) said they were aware of the availability of filter controls they could use; 6 subjects (40%) were somewhat aware and 2 (13.3%) were unaware.

This level of knowledge shifted considerably regarding the awareness of the criteria used to filter messages: only 4 subjects (26.6%) said they were aware of the filter criteria (reliability not tested); 8 subjects (53.3%) were somewhat aware and 3 (20%) were unaware.

The usefulness of the 20 minute session in regards to gaining a deeper understanding of spam filtering issues (1=not at all; 5 =very) was rated slightly better than average (Mean 3.27; Std Dev 1.10) which is not surprising considering the complexity of the topic and the brief period of time.

The improvement of the understanding of what causes false positives and false negatives was medium (Mean 3.33; Std Dev 1.05 and Mean 3.13; Std Dev 1.19, respectively) and could be improved.

However, usefulness of the *spam explanation* and the *advanced spam explanation* were both rated highly (Mean 4.47; Std Dev 0.64 and Mean 4.33; Std Dev 0.98, respectively). The usefulness of the *outgoing mail scanner* (warning users that *outgoing* emails might be classified as spam by the recipient's spam filter) was also rated highly (Mean 3.93; Std Dev 1.07).

CONCLUSIONS AND FUTURE WORK

All of the very popular webmail services discussed in this paper acknowledge and address in some way the problem of what we call spam filter brittleness in previous papers. Regardless they mostly treat spam filtering as a fully automated (black box) process. This means they miss the opportunity to benefit from the fact that spam filters in fact *informate* because they don't use the data generated during message assessments for *informing* (and possibly empowering) users.

An interesting question we raised in earlier papers is whether users of "free" webmail services (and other email services) are actually interested in the details of the spam filtering process works as long as the process appears to be working fine. As demonstrated in this paper there is ample

anecdotal evidence though that spam filters do create certain problems. User-centered design usually suggest IT should be used to empower users in the sense that users should at least have the option of exploring what is happening to their email communication. The feedback we gained from experiments with a simulated email client providing rich information about filtering assessments suggests that making the information available would be appreciated (keeping in mind the audience bias).

It is still unclear though how informing the user about the filtering process could be accomplished as sound technical knowledge may be required to understand the impact of specific filtering criteria on filtering outcomes. We believe technologies used in interactive queries are key to better understanding spam filtering and have started to investigate the incorporation of respective technologies. Dynamic queries are animated user-controlled displays that show information in response to movements of sliders, buttons, maps, or other widgets (eg Ahlberg et al 1992).

Checking not only *incoming* but also *outgoing* email for spamminess is another step towards what Twidale (2004) calls Hubristic Computing by which he means infrastructures (in the broadest sense) both minimizing error and also supporting equivalent effective error recovery by end users. Infrastructures and interfaces explicitly acknowledging their (computational) limitations, allowing for error and supporting error handling/recovery would be extremely beneficial when coping with spam.

ACKNOWLEDGMENTS

The authors wish to thank Michael Twidale for his contributions to establishing this research direction and study participants for their time and effort.

REFERENCES

1. Ahlberg, C., Williamson, C., and Shneiderman, B. (1992), Dynamic queries for information exploration: An implementation and evaluation, Proc. ACM CHI'92: Human Factors in Computing Systems, pp. 619-626.
2. Balvanz, J., Paulsen D. and Struss, J. (2004). Spam software evaluation, training, and support: fighting back to reclaim the email inbox. *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services* Baltimore, MD, USA, 10-13 October 2004. pp. 385-387.
3. Cage, S. (2004). UN seeks spam mandate. *The Australian* 07/07/04. Article available at URL http://australianit.news.com.au/articles/0,7204,10066694%5E15306%5E%5Enbv%5E,00.html
4. CEAS Conference on Email and Anti-Spam 2004-2007. Proceedings (currently up to 2006) available at URL http://www.ceas.cc
5. Fallows, D. (2003). Spam. How it is hurting email and degrading life on the internet. Report published October 22, 2003 by the Pew Internet & American Life project. Washington, DC, USA. © ACM, 2007.
6. Li, X. (2006). E-marketing, unsolicited commercial e-mail, and legal solutions. Webology, 3(1), Article 23. Retrieved October 20, 2006, from http://www.webology.ir/2006/v3n1/a23.html
7. Lueg, C., Huang, J. and Twidale, M. (2007). Mystery Meat revisited: Spam, Anti-Spam Measures and Digital Redlining. *Webology* 4(1) March (ISSN 1735-188X).
8. Lueg, C. (2005). From Spam Filtering to Information Retrieval and Back: Seeking Conceptual Foundations for Spam Filtering. *Proceedings of the 69th Annual Conference of the American Society for Information Science and Technology*, Charlotte NC, USA
9. Lueg, C. (2004). The Hidden Impacts of Anti-Spam Measures and their Contributions to the Digital Divide: An Exploratory Study. *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology,* Providence RI, USA.
10. Martin, S.P. (2006). Informative Spam Interface. BComp (Hons) Thesis, University of Tasmania, Australia.
11. NOIE (2002). Final report of the Australian National Office for the Information Economy (NOIE) review of the spam problem and how it can be countered. Report available at http://www.noie.gov.au/projects/confidence/Improving/Spam/Interim_Report/contents.htm (accessed 03/05/2003).
12. Twidale, M. (2004). Worrying about infrastructures. *Workshop "Distributed Collective Practice: Building new Directions for Infrastructural Studies" of the 2004 ACM Conference on Computer Supported Cooperative Work*, Nov 2004, Chicago.
13. Whitworth, B. and Whitworth, E. (2004). *Spam and the social-technical gap.* IEEE Computer Oct., pp. 38-45.
14. Wright, C. (2003). Can't get spam when you need it. *The Age* November.
15. Zuboff, S. (1988). *In the age of the smart machine*. Basic Books, New York.